AD_____

Award Number: DAMD17-01-1-0787

TITLE: Use of DNA Microarrays to Identify Diagnostic Signature Transcriptional Profiles for Host Responses to Infectious Agents.

PRINCIPAL INVESTIGATOR: J. J. Ellner, M.D.
N. D. Connell, Ph.D.
G. Gallagher, Ph.D.
E. Raveche, Ph.D.

CONTRACTING ORGANIZATION: University of Medicine and Dentistry of New Jersey
New Jersey Medical School
Newark, NJ 07103

REPORT DATE: October 2005

TYPE OF REPORT: Final Addendum

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

| 1. REPORT DATE | 2. REPORT TYPE | 3. DATES COVERED |
|---|---|---|
| 01-10-2005 | Final Addendum | 29 Sep 2004 – 30 Sep 2005 |

**4. TITLE AND SUBTITLE**
Use of DNA Microarrays to Identify Diagnostic Signature Transcriptional Profiles for Host Responses to Infectious Agents.

**5a. CONTRACT NUMBER**

**5b. GRANT NUMBER**
DAMD17-01-1-0787

**5c. PROGRAM ELEMENT NUMBER**

**6. AUTHOR(S)**
J. J. Ellner, M.D.
N. D. Connell, Ph.D.
G. Gallagher, Ph.D.
E. Raveche, Ph.D.

**5d. PROJECT NUMBER**

**5e. TASK NUMBER**

**5f. WORK UNIT NUMBER**

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**
University of Medicine and Dentistry of
New Jersey
New Jersey Medical School
Newark, NJ 07103

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)**
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

**10. SPONSOR/MONITOR'S ACRONYM(S)**

**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

**12. DISTRIBUTION / AVAILABILITY STATEMENT**
Approved for Public Release; Distribution Unlimited

**13. SUPPLEMENTARY NOTES**

**14. ABSTRACT**
Most of the likely agents of bio-terrorism have profound effects on the host and, in particular, on the immune and inflammatory responses. We have proposed a broad-based approach to identify the unique "signatures" of infectious agents using host DNA micro-rrays. Because of the known diverse patterns of host cell interactions with these organisms, examination of the host transcriptional response has enormous potential to allow rapid diagnosis of infectious diseases in general and agents of bio-terrorism in particular. Key signature genes will serve as the basis for rapid diagnostic approaches that could be accessed when an outbreak is suspected. The agents under initial investigation are Bacillus anthracis, Burkholderia mallei, Francisella tularensis, multi-drug resistant Mycobacterium tuberculosis and Yersinia pestis. We have developed appropriate human and mouse models to explore the interactions of host cells with the pathogen. We have begun to characterize the pattern of genes activated or repressed by infection with virulent strains vs. attenuated organisms/mutants as an approach to ensure specificity. Using these data, we hope to identify disease markers to other bio-terrorism agents as well as other common infectious diseases to confirm the specificity of the diagnostic approach.

**15. SUBJECT TERMS**
DAN microarrays; select agents; immune response; pathogenesis

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT | b. ABSTRACT | c. THIS PAGE | | | USAMRMC |
| U | U | U | UU | 15 | 19b. TELEPHONE NUMBER *(include area code)* |

## Table of Contents

Cover.....1

Table of Contents.....2

SF 298.....3

Introduction.....4

Body.....4

Key Research Accomplishments.....11

Reportable Outcomes.....11

Conclusions.....11

References.....N/A

Appendices.....12

## INTRODUCTION

Infections by agents of bioterrorism, especially bacterial agents such as *Bacillus anthracis* and *Yersinia pestis* present their initial symptoms in a way that does not reveal their identity or permit rapid diagnosis. However, as was shown in the recent anthrax attacks on the United States, rapid diagnosis can make the difference between life and death for the patient.

Bacteria, especially virulent bacteria, have a profound effect on the immune system of the human host. Cellular and physiological studies have shown that while many similarities exist in the host immune response to bacterial infection, there are distinctive features that represent classes of organisms and, in some cases, individual organisms themselves. The current proposal was designed to use advanced techniques in molecular analysis to analyze the effect of individual biothreat agents on the human immune system.

We proposed a broad-based approach to identify unique "INFECTION SIGNATURES", characteristic of individual biothreat agents. We used DNA microarrays to undertake these experiments. The high throughput possible and the large scale of investigation (every possible human gene can be examined to determine its expression) has great potential for the diagosis of infectious diseases in general and the presence of biothreat agents in particular. Examination of the gene expression pattern of the human host, in response to *in vitro* infection has the potential to reveal those genes that are regulated by the organism and so reveal its unique characteristics. These can then be exploited; first for diagnosis and later, for therapy and vaccination.

Key "infection signature" genes can serve as the basis for rapid diagnostic approaches that can be used when biothreat attack is suspected. These "infection signature" genes will comprise a set of specific genes that are expressed or repressed during infection by one individual agent, thereby identifying that agent uniquely. This final annual report for Contract # DAMD 17-01-1-0787 summarizes our results with these microarray experiments and presents strong evidence for the existence of these sought-after "infection signatures" for the biothreat agents *Bacillus anthracis, Burkholderia mallei, Francisella tularensis* and *Yersinia pestis*.

## BODY

There were five main aims relevant to the agents under investigation: *Bacillus anthracis, Burkholderia mallei, Francisella tularensis* & *Yersinia pestis*:

1) Develop human and mouse DNA chips to study transcriptional activation and repression by biothreat pathogens.

2) Develop appropriate *in vitro* models to explore the interactions of host cells with the biothreat pathogen and its toxins/constituents.

3) Characterize the pattern of genes expressed or repressed:
    a) by infection with virulent vs attenuated organisms/mutants as an approach to ensure specificity.
    b) in infected cells from vaccinated vs unvaccinated healthy individuals

4) Using the mouse as an *in vivo* model of human infection, characterize changes in gene expression following infection with virulent and avirulent organisms.

5) Based on the above findings, develop DNA chips and assays for associated disease markers that focus on genes and their products that provide the best discrimination between these agents. Apply these DNA chips and assays for disease markers ot other biotreat agents as well as other common infectious diseases to confirm the specificity of the diagnostic approach.

In previous reports we have presented data that address several of these aims.

AIM ONE:
The development of the human and mouse chips was completed quite early and many of our initial data were obtained using chips produced "in-house" for this project. Latterly however, because of change of personnel and the quality problems that ensued, we decided to move over to commercially available microarray chips, choosing the "AFFYMETRIX U133" series chips.

AIM TWO:
In previous reports we have described the use of the "whole blood" infection model. This model closely mimics a natural infection, because the bacteria as used to infect all the cells present in human blood. Often, scientists isolate the lymphocytes (T-cells, B-cells and monocytes) to do such work. We determined that this technique would cause us to lose too much valuable information from the major white blood cell, the neutrophil, and so adopted the technically more demanding whole-blood technique.

AIM THREE:
In previous reports we have presented organism growth curves and host expression analysis that indicates that we have successfully differentiated between infection with virulent vs avirulent strains. We presented an interim analysis of gene expression and showed that early in the infection process with *B.anthracis* (avirulent) there was a marked upregulation of immune system genes, such as chemokines and their receptors, and the cytokines interleukin-6 and tumour necrosis factor. In stark contrast, the virulent strain did not permit the host to activate these elements of the immune system. In the present report (below), we present a more advanced analysis of the gene expression analysis from *B.mallei* infection (the causative agent of "glanders").

Early in the project, we had our source of vaccinated individuals withdrawn and so we have been unable to address this part of Aim three.

AIM FOUR:
The work in the mouse, with *F.tularensis* was extremely successful, as described in the previous report, and is complete. Based on these data, we are pursuing the following continuation studies on the immune mediated effects of infection with F. tularensis:

Based upon studies in this project on mice in which the gene expression profile following infection with F. Tularensis was obtained by microarray analysis, the effects on B cell function specifically has been analyzed. Microarray analysis revealed that many genes related to immunoglobulin production were increased as well as genes downstream of interferons following infection with F. tularensis when compared to uninfected mice profiles. Analysis of sera of mice obtained following F. Tularensis infection demonstrated an overall increase in antibody production, particularly anti-self reactivity. In addition, RT-PCR analysis demonstrated an overall increased expression in interferon lambda following infection with F. Tularensis. In recent experiments, we have found that in vitro exposure of B cell lines to Type I, II and III interferons in some cases increased MHC Class I and II expression. We have constructed plasmid vectors encoding active interferon receptors. We have optimized transfection in both human and mouse B cell lines. Currently we are constructing inducible plasmids for these cytokine receptors and plan to employ them in studies of in vitro infection with F. tularensis. The ultimate goal of these experiments is to alter the gene activation by F. tularensis by altering the cytokine environment with the aim of reducing harmful immune activation while preserving the positive anti-bacterial response

AIM FIVE:
As described below, we have successfully developed a list of genes whose expression or repression indicates infection by one of the four bacteria listed above. Work to develop this list into a viable diagnostic tool is part of the technical memorandum for our allocation under FY'03.

The body of this report then will address results we have obtained that contribute to our successes under Aim 3a. We shall describe gene expression patterns unique to each of the three organisms, which we term their "infection signatures", using methodology described in Appendix One.

**SECTION ONE: defining infection signatures for biothreat agents.**

In previous reports, we have demonstrated successful analysis of gene expression differences between virulent and avirulent strains of *B.anthracis*. In this section, we describe how we have extended that analysis in order to uncover a set of genes whose expression pattern is potentially diagnostic of a particular organism. In brief, the infection signatures were derived to be representative of an organism, irrespective of its virulent status. So, data from the virulent and avirulent strains were combined and compared against the RNA isolated from non-infected (ie cultured only) whole blood. Further, these signatures had to be as widely applicable as possible and so data from all donors were combined, too. In addition, genes whose expression was modulated in response to more than one organism, or all of them, were removed because they would not contribute to the diagnosis of one particular infection. Finally, the *degree* of modulation was considered to be less important that the *fact* of modulation. Thus, we produced a set of genes for each organism, whose expression by the host indicated infection with that organism alone; the "infection signatures. In the following figures, genes illustrated in green are repressed while genes illustrated in red are upregulated.

**1A) Infection signature for *B. anthracis*, the causitive agent of anthrax.**

As shown in Figure One, we defined a clear infection signature for *anthracis*, which comprised 15 downregulated genes (in green) and one upregulated gene (in red). While these genes segregate together for the donors with *anthracis* infection, their expression is completely random for the other

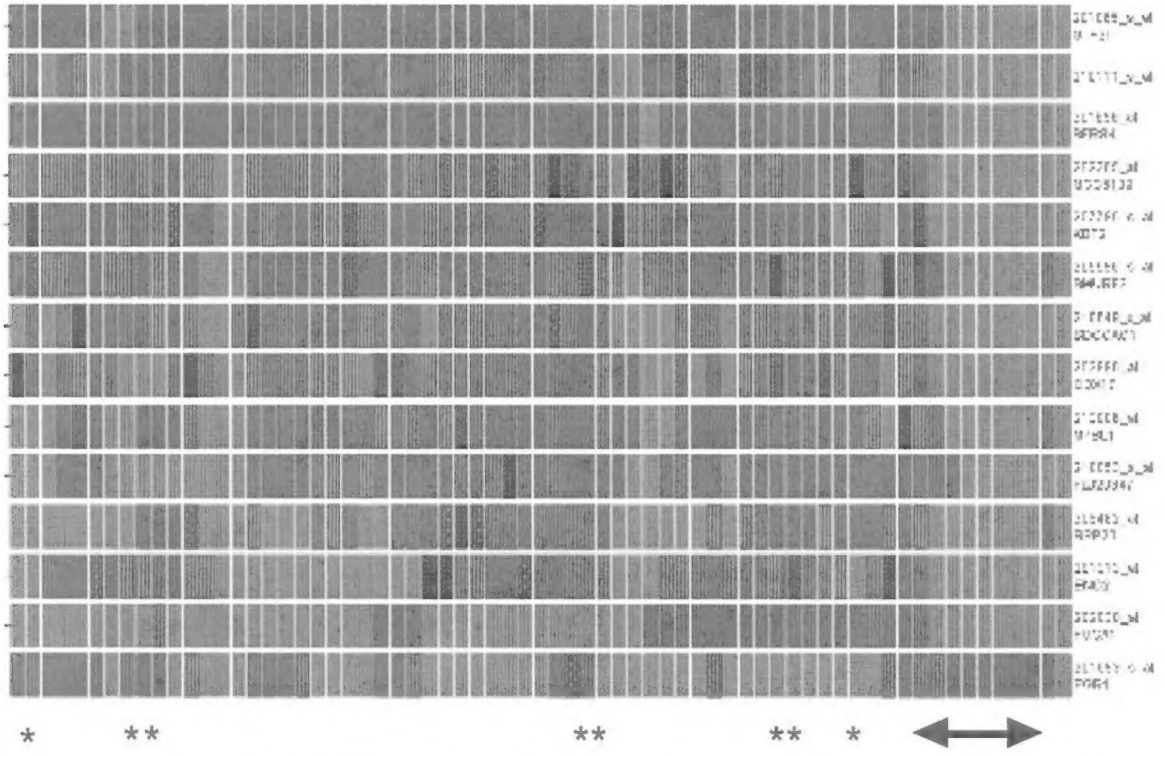


***Figure One: Infection Signature for Anthrax**. The heavy line indicates nine individual donors whose gene expression was similar for the 14 genes indicated on the right of the panel. Asterisks indicate eight donors whose gene expression pattern failed to cluster with the others. The remaining columns represent the same gene expression pattern in the same donors, whose cells were infected with different organisms. Here, a random pattern of gene expression is observed.*

organisms, indicating the specificity of this 'signature' for *anthracis*. The list of genes is shown on the

right (and presented separately in the Appendix Two). With this organism, 9 of 17 donor samples clustered, indicating that 53% of donors showed the same pattern of gene expression. While this is not perfect, it is very encouraging.

Although the cells infected *in vitro* are leukocytes, the genes that make up the infection signature are not immune system genes. This is to be expected because many of these genes would respond similarly to bacterial challenge and so would be screened out by the algorithms used to derive the infection signature.

Thus we have derived a set of gene whose expression in individuals suspected of being infected with a biotreat agent would suggest that the agent was in fact *B.anthracis*.

### 1B) Infection signature for *Y.pestis*, the causative agent of plague.

We analyzed the same data set - comprising all the donors and gene expression for uninfected samples and samples infected with each of the agents - this time asking whether there was a similar gene expression pattern indicative of *Y.pestis* infection. As shown in Figure Two, we identified such a pattern.

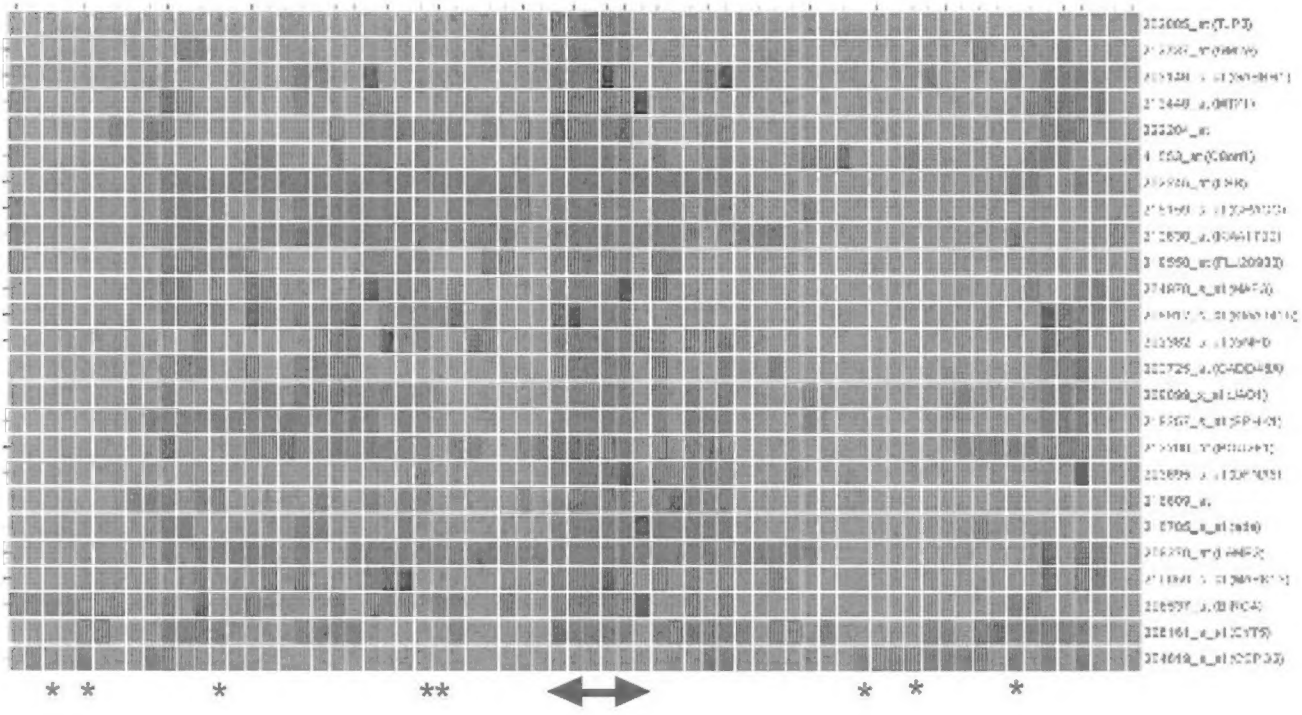


***Figure Two: Infection Signature for Plague**. The heavy line indicates six individual donors whose gene expression was similar for the 25 genes indicated on the right of the panel. Asterisks indicate eight donors whose gene expression pattern failed to cluster with the others. The remaining columns represent the same gene expression pattern in the same donors, whose cells were infected with different organisms. Here, a random pattern of gene expression is observed.*

By the nature of the analysis, this "Plague Infection Signature" contains different host genes from those found in the "Anthrax Infection Signature" shown in Figure One.

In this case there were fourteen donors whose harvested RNA samples were informative in this assay. Six of those clustered to produce a signature of 25 genes while eight fell outside the cluster. Thus, this signature is less informative than the one for Anthrax, since it 'found' only 43% of infected individuals. However, in some ways it is potentially more robust than the Anthrax Infection Signature since 18/25

(72%) of the genes in this signature are upregulated and therefore easier to find, compared with 1/14 (7%) of those inthe Anthrax signature.

Once again, there are no immune response genes present in the signature, as expected, although it is interesting to note that one intracellular signalling gene is included (MAPK13 - a "map" kinase).

### 1C) Infection signature for *B. mallei*, the causitive agent of glanders.

We completed our analysis of select agent infection of human whole blood by searching the dataset for evidence of an infection signature characteristic of *B.mallei* infection. As shown in Figure Three, this was not only present, the "Glanders Infection Signature" was the strongest of the three signatures seen.

In this case there were nineteen donors whose harvested RNA samples were informative in this assay.

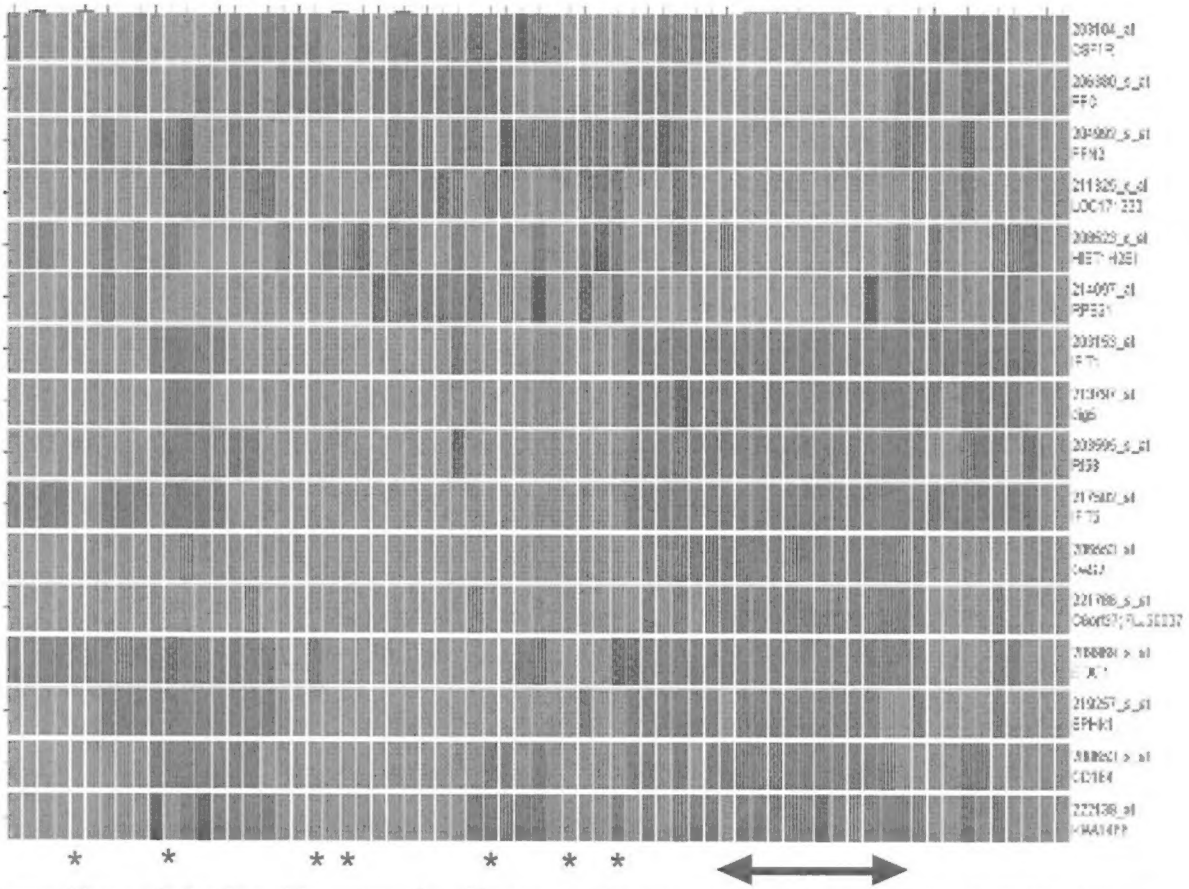

***Figure Three: Infection Signature for Glanders.*** *The heavy line indicates twelve individual donors whose gene expression was similar for the 16 genes indicated on the right of the panel. Asterisks indicate seven donors whose gene expression pattern failed to cluster with the others. The remaining columns represent the same gene expression pattern in the same donors, whose cells were infected with different organisms. Here, a random pattern of gene expression is observed.*

Twelve of those culstered to produce a signature of 16 genes while only seven fell outside the cluster. Thus, this signature is the most informative, since it 'found' 62% of infected individuals. It is equally robust as the Plague Infection Signature, since 50% of the genes in this signature are upregulated.

Two immune response genes are represented in this signature. STAT1 is an intracellular signaling

molecule whose expression is indicative of interferon signalling and usually seen in viral infections. This is in keeping with the intracellular nature of *B.mallei* infection (viruses are intracellular infections also). CD164 is a silaomucin that functions as an "adhesin" molecule. Adhesins are expressed in order for cells to communicate with one another.

**1D) Infection signatures, conclusion.**
We have shown that it is possible to prepare lists of genes (Table One - Appendix) that are modulated in response to specific infections, with approximately 50% efficiency. This is the first time such an observation has been made. The functions of these genes are not related to the specific functioning of the immune system and in some cases these genes have no function yet defined. From the point of our analysis however, this is not important since we are interested in the *fact* of the expression rather than the *function* of the genes expressed.

We intend to take this knowledge forward to produce advanced diagnostic tools which will indicate the presence of these agents within individuals, while disease is asymptomatic. Furthermore, these tools. may be capable of use as a screening tool for prophylaxix - individuals with no evidence of exposure may be spared extensive antibiotic use, for example.

**SECTION TWO: understanding pathways of pathogenesus using expression analysis.**

In addition to the infection signatures described above, we have begun to investigate the mechanisms of pathogenesis used by bioweapon agents. In this example, we describe an analysis of differences seen between two strains of *B.mallei*, when they infect human cells. We have used the **virulent** (ATCC 23344/China 7) and **avirulent** (85-503) strains of *B. mallei*, and provide evidence for possible mechanisms of virulence and host immune evasion.

The bacterium naturally infects horses, mules and donkeys, but can be acquired by humans principally through handling infected animals or laboratory specimens. There is no vaccine against glanders, and although many strains are susceptible to antibiotics, including penicillins and tetracyclines, there is still a 95% fatality rate for untreated septicemia, (50% with treatment) and an overall fatality rate of 40%for all routes of infection. *B. mallei* is easily aerosolized and highly infectious, thus making it a desirable agent of bioterrorism. It has previously been used as a bioweapon, targeting the equine population on several occasions, including WW1, in an effort to incapacitate the cavalry and prevent advancement of an opposing army.

The pathogenesis of *B.mallei* is not well characterized, but potential mechanisms of the pathogenesis include a polysaccharide capsule, numerous secreted proteases and a type III secretion system (TTSS). While recent work has uncovered several mechanisms by which B. mallei may propagate disease, little is known about the host response during infection. In this study, an in vitro whole blood infection model was used to gather information about the response to both **virulent** (ATCC 23344/China 7) and **avirulent** (85-503) strains of *B. mallei*.

We made use of the "PathwayAssist" software package to explore the likely biological role, and interactions, of gehes shwon to be expressed in the microarray experiments involvong these two strains. "PathwayAssist" performs this task by comparing experimental data to curated data in the ResNet mammalian database. This database is generated using the MedScan text mining technology, which scans sentences from Medline and PubMed abstracts to extract useful representations of the interactions described in the literature. The result is a database describing over 500,000 biological interactions between proteins, chemical compounds, small molecule therapeutics etc. These interactions combine to form an anastomosing network that assists in uncovering common pathways between seemingly unrelated genes.

The data illustrated in Figure Four represent the analysis of host mRNA from 8 hour virulent vs 8 hour avirulent samples. The plasma membrane and nucleus have been included to convey spatial relationships

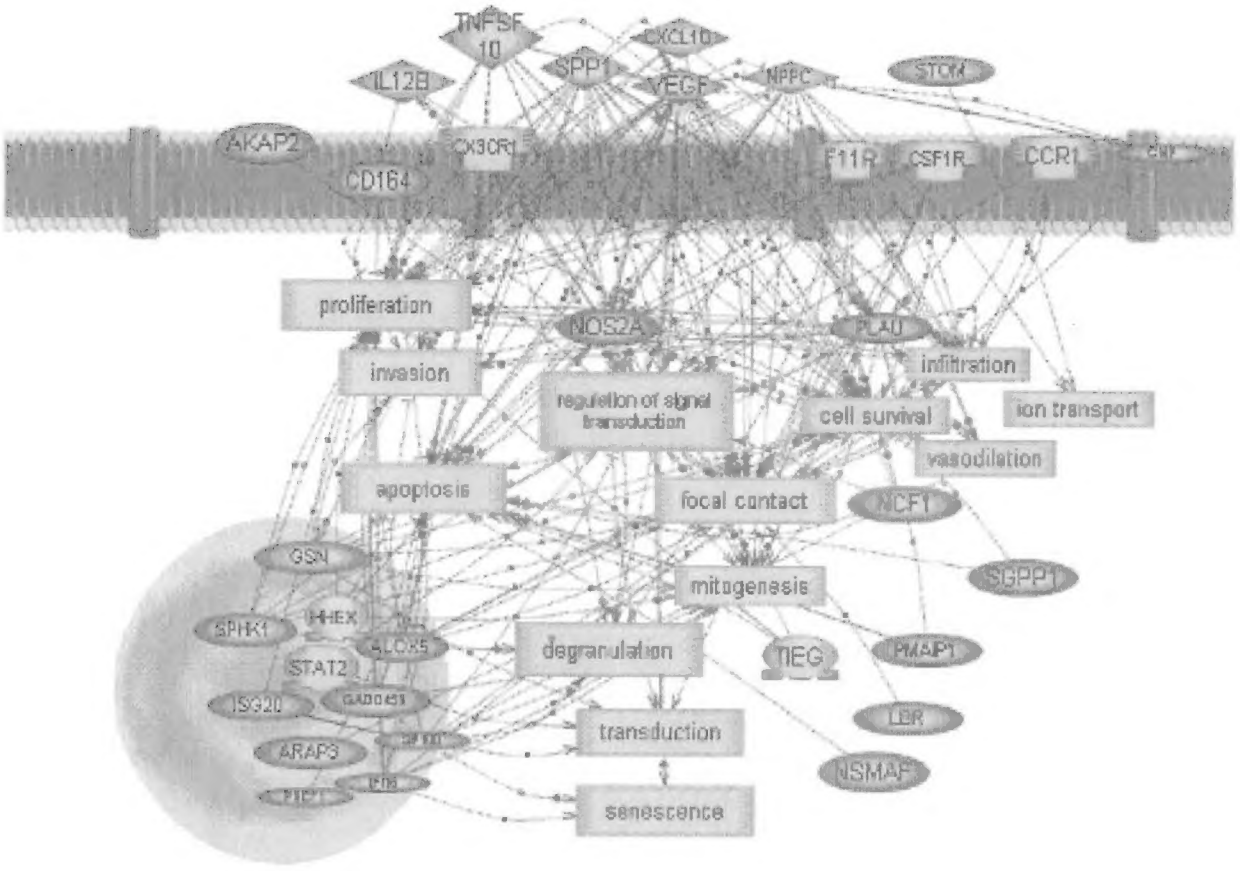


***Figure Four: Expression analysis for Glanders.*** *The figure represents the comparative analysis of host response to virulent vs avirulent glanders, at 8 hours into the infection. Proteins are shown in red and cellular processas are shown in yellow. The cell membrane and nucleus are shown for orientation.*

between interacting proteins (shown in red, while cellular processes shown in yellow). Notable processes involved after 8 hour infections include apoptosis, cell survival, focal contact and regulation of signal transduction. In addition to these processes, the avirulent strains induce such processes as motility, proliferation and secretion.

Both China 7 (virulent) and 85-503 (avirulent) each induce divergent changes in the expression patterns of genes involved in signal transduction/cell signaling. China 7 downregulates several genes that are critical to the developing immune responses, such as colony stimulating factor-1 receptor (CSF1R), chemokine C-X3-C motif receptor 1 (CX3CR1) Secreted Phosphoprotein 1 (SPP1) and nuclear cytosolic factor 1 (NCF1). **CSF1R** plays a major role in the antimicrobial humoral response and development and function of monocytes and macrophages, and **CX3CR1** stimulates actin polymerization and chemotaxis in various cells. **SPP1**, otherwise known as osteospontin, positively regulates T-cell proliferation and is involved in the T-helper type 1 immune response. **NCF1** (neutrophil cytosolic factor 1) is involved in the cellular response to bacteria via electron transport and superoxide metabolism. One final cellular process that is differentially affected is the rearrangement of the cytoskeleton. **Gelsolin (GSN)** is downregulated upon infection with China 7, and controls actin filament polymerization through both capping and severing activity. The gelsolin superfamily of proteins have been implicated in numerous processes, including cell motility and regulation of phagocytosis. GSN represents another possible target for China 7 to disrupt, as downregulation of this gene potentially hinders phagocytosis bacteria and formation of phagolysosomes. Moreover, the inhibition of actin depolymerases may contribute to the intracellular spread of B. mallei, as

similar organisms induce the formation of long, cellular protrusions which facilitate infection of neighboring cells. Despite inducing a seemingly global reduction in immune responses, China 7 also induces expression of genes involved in cell signaling, suggesting these genes represent key components of pathways that are not interrupted by the bacteria.

Of greater interest are the genes that are induced upon stimulation with 85-503 but not China 7. These genes again represent key differences in the immune response to each strain, and offer insight into the pathogenicity of China 7. A list of these genes is shown in APpendix Three.

**CD69** is one such receptor, and is rapidly expressed on activated T cells, B and NK cells, but not resting lymphocytes. CD69 is also present on activated macrophages and other cell types, including neutrophils, eosinophils and platelets, and mediates cytokine production through calcium signaling. Another receptor upregulated by 85-503 stimulation is **CCRL2**. This gene closely resembles the chemokine receptor CCR1, and is expressed at high levels in primary neutrophils and primary monocytes, and is further upregulated on neutrophil activation and during monocyte to macrophage differentiation. Though its function is not completely understood, it is potentially involved in chemotaxis propagation of the innate immune response. **TRIP10**, a gene upregulated upon infection with 85-503, is a microtubule binding protein involved in the regulation of cell shape and polarity. It has a main role in actin cytoskeleton reorganization and biogenesis. It also induces podosome formation in macrophages, increasing its interactions with surrounding cells and the local environment.

**Key research accomplishments:**

Microarray detection of gene expression of *B. anthracis, Y. pestis, B. mallei, M. tuberculosis* and SARS have been completed

Microarray detection of gene expression of *F. tularensis*, Hanta virus, Dengue virus, Influenza virus and monkeypox virus are near completion.

Algorithms have been developed and used successfully to extract infection signatures for three of the ten infectious agents we are studying.

**Reportable outcomes:** Manuscripts in preparation.

**Conclusions.**

**1. Infection signatures**

We have shown that it is possible to prepare lists of genes (Table One - Appendix) that are modulated in response to specific infections, with approximately 50% efficiency. This is the first time such an observation has been made. The functions of these genes are not related to the specific functioning of the immune system and in some cases these genes have no function yet defined. From the point of our analysis however, this is not important since we are interested in the *fact* of the expression rather than the *function* of the genes expressed. *We intend to take this knowledge forward to produce advanced diagnostic tools which will indicate the presence of these agents within individuals, while disease is asymptomatic.* Furthermore, these tools may be capable of use as a screening tool for prophylaxis - individuals with no evidence of exposure may be spared extensive antibiotic use, for example.

**2. Expression analysis**

The disparity between host responses to virulent and avirulent strains of *B. mallei* suggests there are pivotal differences in their ability to subvert the immune response and propagate infection. Here we describe, for the first time, distinct pathways involved in immune responses, signaling, and cell adhesion/cytoskeletal rearrangement, that will further the understanding of China 7 pathogenesis. Not only will this knowledge help to design more effective therapeutics, but will aid in characterizing the novel virulence factors utilized by China 7.

## APPENDIX ONE: Experimental Protocols for Microarray analysis and Infection Signature Generation

### Sample Processing

Double-stranded cDNA was synthesized from total RNA using the Superscript Double Stranded Synthesis kit (Invitrogen). Following phenol/chloroform extraction and ethanol precipitation, a biotin-labeled *in vitro* transcription reaction was carried out using the cDNA template (Enzo Bioarray). The resultanat cRNA (15 °g) was fragmented and added to a hybridization mixture containing control cRNA and oligonucleotides. Hybridization proceeded at 45∞C with rotation for 16 hours using the Genechip Hybridization Oven 640 (Affymetrix). Washing and staining (Streptavidin Phycoerythrin) was performed using the Genechip Fluidics Station 400 (Affymetrix) using the EukGE-WS2v4 protocol. Images were acquired using the Affymetrix GeneArray scanner. Data was extracted using Affymetrix Microarray Suite 5.0. The quality of each chip was determined and chips were used in the data analysis if they met the following criteria: The background was less than 100, the noise was in the range of 1-3 and the GAPDH 3' to 5' ration was between 1 and 3. Detailed protocols can be found in the file called CAG Affy Protocols and the Affymetrix Desktop GeneChip Expression Manual included with this report.

### Data Analysis

**GeneChip Data Scaling:** All GeneChips were scaled to a mean target intensity of 500 to allow comparisons between arrays within a data set. The intensity values of all genes on a chip were calculated and the top and bottom 2% were removed. A scaling factor was computed by dividing the target mean by the mean of the remaining genes. The chip was then scaled by multiplying all of the genes by this scaling factor.

**Determination of present/absent calls:** Present/absent calls for all transcripts on the array were made using Affymetrix MAS5. The detection algorithm uses probe pair intensities to generate detection p-values. Using the default settings, a probe set with a detection p-value less than 0.4 is considered present (represented by a letter 'P' in the detection column of any given sample), a p-value between 0.6 and 0.4 is considered marginal (represented by a letter 'M'), and a p-value greater than 0.6 is considered absent (represented by a letter 'A'). A weight of reliability can be based on the specific p-value, i.e., a p-value closer to 0 indicates a gene more likely to be present. On the other hand, a p-value closer to 1 indicates a gene more likely to be absent. Details of the Affymetrix detection algorithm can be found in the statistical reference guide included with this report.

**Fold change calculations:** Log ratios and change calls were calculated using Affymetrix MAS5. Each probe set on the experiment array is compared to its counterpart on the control array, and a change value is calculated indicating an increase, decrease or no change in gene expression. This change is expressed as the log2 ratio. The signal log ratio is related to the fold change by the following formulas:
Fold change = 2signal Log ratio for signal log ratio≥0
Fold change = (-1) x 2- signal log ratio for signal log ratio<0
Change calls are designated as Increase, Marginal Increase, No Change, Decrease, or Marginal Decrease. Details of the change Affymetrix algorithm and terminology can be found in the statistical reference guide.

**Gene list creation:** Gene lists were created using Microsoft Access and Microsoft Excel. The data generated with MAS5 were loaded into Microsoft Access. Gene lists were formed from queries based upon the criterion of the test being used, ie. Absent/present call and fold changes. The data from the queries were imported to Microsoft Excel.

**Gene Selection based on change calls:** For a given condition (1st condition), a gene was classified as up regulated if in that sample the gene was called present, there was an increase call between that the 1st condition and uninfected, and the 2nd condition must not have had a larger response than the 1st condition. Similarly, for a gene to be classified as down regulated, the gene was called present in the uninfected sample, the 1st condition to uninfected was called a decrease, and the 2nd condition must not

have had a smaller response than the 1st condition. For example, to determine genes up-regulated in the virulent sample compared to the avirulent and uninfected samples, the gene was called present in the virulent sample, the virulent vs uninfected comparison called an increase and virulent to avirulent comparison did not make a decrease call.

**Gene Selection based on present/absent calls:** This method uses the present/absent calls made by MAS5 but does not take into account log ratio or fold change calculations. For any given condition, the genes that were present in all of the chips or absent in all of the chips for that condition were identified. A gene was considered significant if it was present in that condition and absent in the other two conditions or absent in that condition and present in the other two. For example, a gene was considered significant if it was called present in the virulent condition and absent in the corresponding uninfected and avirulent conditions.

**Selective expression analysis:** Selective expression analysis also uses the present/absent calls made by MAS5 and does not use log ratio or fold change calculations. In the selective expression algorithm, each gene is assigned a value based on its present call, 1 for present and 0 for absent. The selectivity of each gene between two groups is the mean of the absolute difference of the gene's value between the two groups. Genes with higher scores are more likely to be consistently differentially expressed. By shuffling the data, we can distinguish whether the grouping was random or significant. This algorithm was developed using data from acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) patients (Golub et al., 1999, Science 286:521-537). We have successfully used it to classify all the samples in the training set and the 33 out of the 34 samples in the independent set of samples. The selective expression method is robust and well suited to identify diagnostic gene markers.

**Cluster analysis:** Cluster analyses were performed using GeneSpring (Silicon Genetics). Hierarchical clustering was performed on the genes based on expression levels across samples. Genes that display similar expression patterns can be visualized based on where they appear on the branches of the tree. Clustering determines the similarity between two or more samples or genes by applying a similarity metric. Samples and genes that have similar expression patterns cluster together and those that are dissimilar cluster far apart.

**Annotations:** All gene annotations were obtained from the NetAffx Analysis Center on the Affymetrix Website (http://www.affymetrix.com/analysis/index.affx) or by using GeneSpider in the GeneSpring package.

**APPENDIX TWO: PCR-verified genes to be utilized in the "Infection Signatures"**

**ANTHRAX**

| Gene Title | Genbank # |
|---|---|
| enolase 2, (gamma, neuronal) | NM_001975 |
| early growth response 1 | NM_001964 |
| splicing factor, arginine/serine-rich 4 | NM_005626 |
| hypothetical protein MGC5139 | BC004815 |
| fucosidase, alpha-L- 1, tissue | NM_000147 |
| popeye domain containing 2 | NM_005694 |
| E3 ubiquitin ligase SMURF2 | NM_022739 |
| POP7 (processing of precursor, S. cerevisiae) homolog | NM_005837 |
| KIAA0265 protein | NM_014997 |
| KIAA0853 protein | NM_015070 |
| v-myb myeloblastosis viral oncogene homolog (avian)-like 1 | X66087 |
| neuroendocrine differentiation factor | NM_016079 |
| serologically defined colon cancer antigen 1 | NM_004713 |
| Notch homolog 1, translocation-associated (Drosophila) | NM_017617 |
| hypothetical protein FLJ20847 | XM_373413 |
| hypothetical protein FLJ20094 | NM_017665 |

**GLANDERS**

| Gene Title | Genbank # |
|---|---|
| 2',5'-oligoadenylate synthetase 1, 40/46kDa | NM_016816 |
| Interferon-induced protein with tetratricopeptide repeats 1 | NM_001548 |
| Interferon-induced protein with tetratricopeptide repeats 5 | NM_012420 |
| chromosome 1 open reading frame 29 | NM_006820 |
| 2',5'-oligoadenylate synthetase 1, 40/46kDa | NM_002534 |
| properdin P factor, complement | NM_002621 |
| 2'-5'-oligoadenylate synthetase 2, 69/71kDa | NM_002535.1 |
| CD164 antigen, sialomucin NM_006016.3 | |
| signal transducer and activator of transcription 1, 91kDa | NM_139266 |
| viperin | NM_080657 |
| interferon-induced protein with tetratricopeptide repeats 2 | NM_001547 |
| sphingosine kinase 1 | NM_021972 |
| hypothetical protein FLJ20637 | NM_017912 |
| chromosome 6 open reading frame 37 | AF350451 |
| KIAA1466 protein | AB040899 |

**PLAGUE**

| Gene Title | Genbank # |
|---|---|
| tight junction protein 2 (zona occludens 2) | NM_001005752 |
| deafness, autosomal dominant 5 | NM_004004 |
| growth arrest and DNA-damage-inducible, alpha | NM_001007468 |
| synaptotagmin V | NM_000719 |
| laminin, beta 3 | NM_024426 |
| mitogen-activated protein kinase 13 | NM_002754 |
| phospholipase A2, group IVA (cytosolic, calcium-dependent) | NM_008869 |
| GM2 ganglioside activator protein | NM_000405 |
| transcribed sequence (H.Sapiens) | NP_062553.1 |
| hypothetical protein FLJ11267 [Homo sapiens] | BC017979 |

**APPENDIX THREE: Host genes upregulated in repsponse to virulent *B.mallei*, at eight hours of infection.**

| 8 Hour Fold Change | Genbank | Name |
|---|---|---|
| 10.68 | NM_017912 | hypothetical protein FLJ20637 |
| 8.55 | AL576654 | ribosomal protein S20 |
| 6.71 | AL576654 | ribosomal protein S20 |
| 5.85 | NM_019618 | interleukin 1 family, member 9 |
| 4.85 | AF022375.1 | vascular endothelial growth factor |
| 4.08 | NM_014398 | lysosomal-associated membrane protein 3 |
| 4.00 | AF078077.1 | Growth arrest and DNA-damage inductible gene 45 Beta |
| 3.76 | NM_002201 | interferon stimulated gene 20kDa |
| 3.50 | NM_002658 | plasminogen activator, urokinase |
| 3.46 | NM_021972 | sphingosine kinase 1 |
| 3.38 | NM_001565 | chemokine (C-X-C motif) ligand 10 |
| 2.66 | AF208043.1 | interferon inducible protein 16 |
| 2.58 | NM_005419 | signal transducer and activator of transcription 2, 113kDa |
| 2.52 | BF575514 | pre-B-cell colony-enhancing factor |
| 2.49 | NM_002187 | interleukin 12B, p40 |
| 2.45 | NM_018381 | hypothetical protein FLJ11286 |
| 2.44 | AB002319 | KIAA0321 protein |
| 2.41 | NM_001295 | chemokine (C-C motif) receptor 1 |
| 2.38 | AI857639 | phorbol-12-myristate-13-acetate-induced protein 1 |
| 2.35 | NM_014597 | acidic 82 kDa protein mRNA |
| 2.35 | NM_005655 | TGFB inducible early growth response |
| 2.29 | AI421071 | chemokine (C-C motif) receptor 1 |
| 2.23 | AI537887 | stomatin |
| 2.22 | NM_003113 | nuclear antigen Sp100 |
| 2.11 | AU134977 | Human clone 137308 mRNA, partial cds. |
| 2.10 | D79994.1 | Kidney Ankyrin Repeat Containing Protein |
| 2.10 | NM_003580 | neutral sphingomyelinase (N-SMase) activation associated factor |
| 2.09 | U57059.1 | Tumor Necrosis Factor superfamily 10 |
| 2.08 | AB007447 | FLN29 gene product |
| 2.07 | NM_024989 | hypothetical protein FLJ12377 |
| 2.03 | BC001362.1 | C-type natriuretic peptide |
| 2.02 | AW129593 | tudor repeat associator with PCTAIRE 2 |